

Moissonnage de catalogues Open Data

client : Région Hauts-de-France

Travaux réalisés en 2021 – Présentation au CNIG le 28/03/2022
Version 1.2

Contexte, objectifs

Parties prenantes

Région Hauts-de-France pour Géo2France

Neogeo Technologies
(interventions de Benjamin Chartier en sous-traitance)

Camptocamp

Géo2France

Plateforme régionale partenariale :
État/Région avec implication d'autres acteurs locaux

Portail internet : geOrchestra + d'autres outils

Catalogage :

- **GeoNetwork** pour le catalogage de **ressources géographiques**
- **CKAN** : pour le catalogage de **données ouvertes**
- Mécanismes de moissonnage pour récupérer des métadonnées dans des catalogues externes
- Mécanismes de moissonnage pour alimenter CKAN à partir de GeoNetwork

Objectifs des travaux de 2021

Proposer un point d'accès unique aux données sur le territoire régional :
DataHub régional

Simplifier l'accès aux données

- Un seul catalogue pour l'ensemble des métadonnées (ne plus avoir le doublon CKAN - GeoNetwork)
- Simplification de l'interface de recherche et de découverte du patrimoine de données
- Mieux présenter les métadonnées
- Mieux gérer les liens entre métadonnées
- Présenter un aperçu des données

Moissonner la plupart des catalogues open data des partenaires locaux
[Isogéo, OpenDataSoft, ArcGIS Online, uData (data.gouv.fr), CKAN]

Passer de la version 3 à la version 4 de GeoNetwork

Une certaine biodiversité
[à l'échelon régional mais aussi à plus large échelle]

**Solutions de catalogage
très variées**

CKAN

DKAN

Isogéo

GeoNetwork

uData (data.gouv.fr)

OpenDataSoft

ArcGIS Online

[Géoclip]

[Géocatalogue]

Protocoles et formats très variés

DCAT et ses dérivés

ISO 19115 (versions 2005 et 2014)

API :

CSW,

Autres API spécifiques

[ArcGIS, OpenDataSoft, CKAN, uData...]

OGC API – records

DCAT – quoi ça ?

DCAT : Data Catalog Vocabulary

Standard du W3C (ouvert et gratuit)

Vocabulaire RDF :

- Modèle de graphe qui permet de décrire des ressources du web et d'établir des relations entre elles (web sémantique).
- Élément de base de RDF : triplet sujet-prédicat-objet
- **Ni un protocole d'interrogation d'un catalogue ni un format de données**
- Peut être retranscrit selon **plusieurs formats** : RDF/XML, Turtle, JSON-LD, RDFa par exemple

À noter :

- Peu de contraintes sur le renseignement des métadonnées (peu d'informations requises, pas de vocabulaires externes)
- Principaux concepts de DCAT : **Catalog, Dataset, Distribution**

Exemple DCAT v2 multilingue avec une bbox [au format Turtle]

```
<Dutch-postal> a dcat:Dataset ;

    dct:title "Adressen"@nl ;
    dct:title "Addresses"@en ;

    dct:description ""INSPIRE Adressen afkomstig uit de
basisregistratie Adressen, beschikbaar voor heel Nederland""@nl ;
    dct:description ""INSPIRE addresses derived from the Addresses
base registry, available for the Netherlands""@en ;

    dcat:theme <http://inspire.ec.europa.eu/theme/ad> ;

    dct:spatial [
        a dct:Location ;
        dcat:bbox ""POLYGON((
            3.053 47.975 , 7.24 47.975 ,
            7.24 53.504 , 3.053 53.504 ,
            3.053 47.975
        ))""^^geosparql:asWKT ;
    ] .
```

cf. <https://www.w3.org/TR/vocab-dcat-2/#ex-spatial-coverage-bbox>

Famille DCAT

3 versions : version 2 publiée en 2020, version 3 en cours d'élaboration

Plusieurs déclinaisons au niveau européen :

- DCAT-AP : profil défini pour s'adapter au contexte réglementaire local (directive PSI).
- StatDCAT-AP : spécialisation pour les données statistiques
- GeoDCAT-AP : même chose pour les données géolocalisées.

Objectif : faire rentrer des métadonnées INSPIRE dans DCAT (aux forceps).

Attention ce travail date de 2016 (antérieur à la DCAT v2!)

Un brouillon de décembre 2020 basé sur DCAT v2 existe néanmoins.

ISO 19115

2 versions majeures :

Version 2005 utilisée très majoritairement en France dans le contexte INSPIRE

Version 2014 plus rarement utilisée (utilisé en Wallonie par exemple)

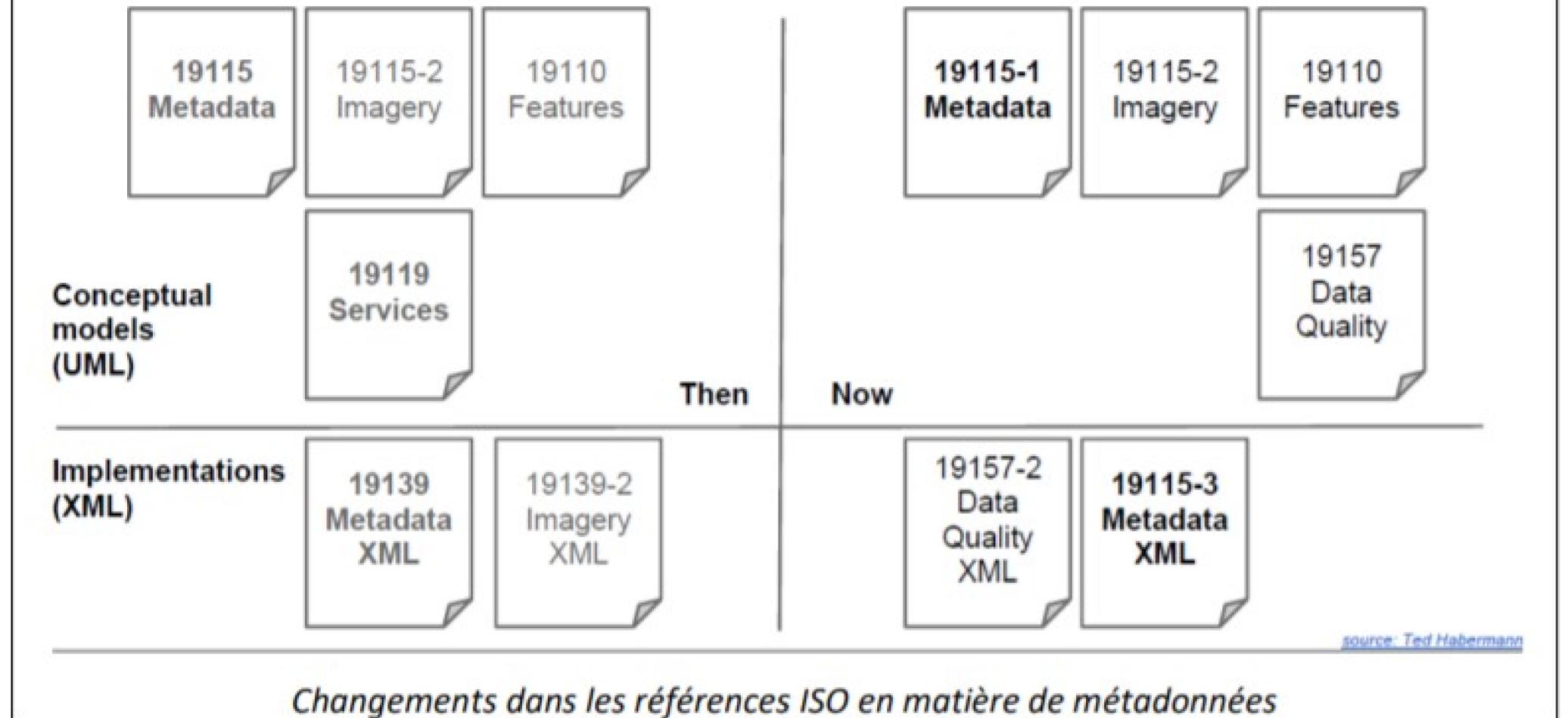
Standard de l'ISO (**payant et très cher !**)

ISO 19115

2005

2014

“I am confused by all these numbers!”



ISO 19115-1:2014

Quelques évolutions notables (voir annexe G du standard) :

Ce standard couvre les **métadonnées de données et de services**

Qualité transféré dans ISO 19157

Possibilité de décrire le **format de chaque distribution**
(MD_DigitalTransferOptions/distributionFormat: MD_Format)

Possibilité d'intégrer un **dictionnaire d'attributs**
(MD_FeatureCatalogue)

Nouveau champ pour l'**adresse en ligne de la fiche de métadonnées**
(MD_Metadata/metadataLinkage:CI_OnlineResource)

Nouveau champ pour le **référencement les métadonnées dans des formats alternatifs** (MD_Metadata/alternativeMetadataReference)

Nouveau champ pour le **référencement d'une documentation supplémentaire** (MD_Identification/additionalDocumentation: CI_Citation)

Quelques différences entre DCAT et ISO 19115

Articulation dataset / distribution :

- Pour ISO 19115, quasiment aucune information portée par les distributions
- Pour DCAT, beaucoup plus d'informations au niveau des distributions :
 - format (maintenant aussi pour ISO 19115-1:2014)
 - licence
 - résolution spatiale

Mots-clefs / thèmes :

- Notion de thème pour DCAT (largement utilisée par les catalogues open data) en plus de celle des mots-clefs. Ces thèmes sont souvent spécifiques à chaque portail open data.
- Pour ISO 19115 on a un savant mélange de mots-clefs et de catégories thématiques (qui ne correspondent pas aux thèmes des portails open data).

En bref

L'interopérabilité / le moissonnage de métadonnées :

Toujours pas simple

De plus en plus compliqué

Quelles métadonnées exposées
par les solutions Open Data ?

ArcGIS Online – DCAT - JSONLD

```
{
  "@type": "dcat:Dataset",
  "identifiant": "https://data-atmo-hdf.opendata.arcgis.com/datasets/atmo-hdf::ind-hdf-agglo",
  "landingPage": "https://data-atmo-hdf.opendata.arcgis.com/datasets/atmo-hdf::ind-hdf-agglo",
  "title": "ind hdf agglo",
  "description": "Tableau de données et visualisation de l'indice de la qualité de l'air en Hauts-de-France sur 1 an glissant. <div><br /></div><div>L'indice de qualité de l'air est un indicateur chiffré [ ... ]</div>",
  "keyword": ["indice", "agglomeration", "visualisation"],
  "issued": "2019-09-25T08:28:30.000Z",
  "modified": "2021-01-26T16:35:04.799Z",
  "publisher": {"name": "Atmo Hauts-de-France" },
  "contactPoint": {
    "@type": "vcard:Contact",
    "fn": "AtmoHDF",
    "hasEmail": "mailto:contact@atmo-hdf.fr"
  },
  "accessLevel": "public",
  "spatial": "1.6099,49.2557,3.9613,51.0292",
  "license": "http://opendatacommons.org/licenses/odbl/summary",
  "distribution": [
    {
      "@type": "dcat:Distribution",
      "title": "GeoJSON",
      "format": "GeoJSON",
      "mediaType": "application/vnd.geo+json",
      "accessURL": "https://data-atmo-hdf.opendata.arcgis.com/datasets/atmo-hdf::ind-hdf-agglo.geojson?[ ... ]"
    },[ ... ]
  ],
  "theme": ["geospatial"]
}
```

OpenDataSoft – JSON – partie 1

```
{
  "datasetid": "courbe-consommation-agreee-inf-36kva-par-region",
  "metas": {
    "domain": "agenceore",
    "staged": false,
    "visibility": "domain",
    "metadata_processed": "2022-02-09T15:00:58.058147+00:00",
    "data_processed": "2022-02-09T15:00:58+00:00",
    "modified": "2022-02-09T14:10:37+00:00",
    "license": "Licence Ouverte (Etalab V2)",
    "publisher": "Agence ORE & Gestionnaires de réseaux électricité et gaz",
    "description": "<p>Ce jeu de données présente une agrégation de la consommation [ ... ].</p>",
    "theme": ["Consommation"],
    "title": "Consommation agrégée demi-horaire des points de soutirage inférieurs à 36kVA par région",
    "keyword": ["Article 23"],
    "license_url": "https://www.etalab.gouv.fr/wp-content/uploads/2017/04/ETALAB-Licence-Ouverte-v2.0.pdf",
    [ ... ]
    "modified_updates_on_data_change": false,
    "metadata_languages": ["fr"],
    "geographic_reference": ["world_fr"],
    "language": "fr",
    "territory": ["France"],
    "records_count": 526176,
    "frequence-de-mise-a-jour": "Trimestrielle",
    "maille-geographique": "Région",
    "energie": ["Électricité"],
    "pas-temporel": "1/2 Horaire",
    "thematique": ["Consommation"],
    "profondeur-dhistorique": "2020 - 2021",
    "reseaux": ["Distribution"]
  }
}
```

OpenDataSoft – JSON – partie 2

```
"has_records": true,  
"data_visible": true,  
"features": [  
  "timeserie",  
  "analyze"  
],  
"attachments": [  
  {  
    "mimetype": "application/pdf",  
    "id": "note_methodologique_agregats_de_conso_demi_horaire_des_points_inf_a_36kva_par_region_v2_0_pdf",  
    "title": "Note methodologique - Agrégats de conso demi-horaire des points inf à 36kVA par région v2.0.pdf",  
    "url": "odsfile://note-methodologique-agregats-de-conso-demi-horaire-des-points-inf-a-36kva-par-region-v20.pdf"  
  }  
],  
"alternative_exports": [],  
"fields": [  
  {  
    "name": "horodate",  
    "description": "Horodate de la mesure",  
    "annotations": [  
      {  
        "name": "facet"  
      }  
    ],  
    "label": "HORODATE",  
    "type": "datetime"  
  },  
  [ ... ]  
]
```

uData – DCAT

```
{
  "@id": "https://www.data.gouv.fr/datasets/62384eaba33290ad418cb3f0/",
  "@type": "Dataset",
  "description": "La Cour publie des notes faisant partie d'un ensemble de travaux menés sur plusieurs grandes politiques publiques, qui identifient à la fois les principaux défis auxquels seront confrontés les décideurs publics au cours des prochaines années, et les leviers susceptibles de les relever. Dans le prolongement du rapport « Une stratégie de finances publiques pour la sortie de crise », remis en juin dernier au Président de la République et au Premier ministre, cette série de publications permet à la Cour de s'exprimer sur des sujets structurels. Avec ces notes à visée pédagogique, la Cour se place au cœur de sa mission d'information des citoyens, à un moment démocratique clé.",
  "distribution": "https://www.data.gouv.fr/datasets/62384eaba33290ad418cb3f0/#resource-c48182e2-a124-4302-899a-de6a92ebc976",
  "identifiant": "62384eaba33290ad418cb3f0",
  "issued": "2022-03-21T11:08:43.571000",
  "keyword": [
    "enseignement-superieur",
    "universites"
  ],
  "modified": "2022-03-21T11:09:06.747000",
  "publisher": "https://www.data.gouv.fr/organizations/53698dada3a729239d20331d/",
  "title": "Les universités à l'horizon 2030 : plus de libertés, plus de responsabilités"
},
```

uData – API REST – JSON – partie 1

```
{
  "acronym": null,
  "archived": null,
  "badges": [],
  "created_at": "2022-03-21T11:08:43.571000",
  "deleted": null,
  "description": "La Cour publie des notes faisant partie d'un ensemble de travaux menés sur plusieurs grandes politiques publiques, qui identifient à la fois les principaux défis auxquels seront confrontés les décideurs publics au cours des prochaines années, et les leviers susceptibles de les relever. Dans le prolongement du rapport « Une stratégie de finances publiques pour la sortie de crise », remis en juin dernier au Président de la République et au Premier ministre, cette série de publications permet à la Cour de s'exprimer sur des sujets structurels. Avec ces notes à visée pédagogique, la Cour se place au cœur de sa mission d'information des citoyens, à un moment démocratique clé.",
  "extras": {},
  "frequency": "unknown",
  "frequency_date": null,
  "id": "62384eaba33290ad418cb3f0",
  "last_modified": "2022-03-21T11:09:06.747000",
  "last_update": "2022-03-21T11:09:06.694000",
  "license": "odc-odbl",
  "metrics": {[ ... ]},
  "organization": {
    "acronym": null,
    "badges": [ ... ],
    "class": "Organization",
    "id": "53698dada3a729239d20331d",
    "logo": "https://static.data.gouv.fr/avatars/ac/83f761a23f4263be8b3401b01625e8-original.png",
    "logo_thumbnail": "https://static.data.gouv.fr/avatars/ac/83f761a23f4263be8b3401b01625e8-100.png",
    "name": "Cour des comptes",
    "page": "https://www.data.gouv.fr/fr/organizations/cour-des-comptes/",
    "slug": "cour-des-comptes",
    "uri": "https://www.data.gouv.fr/api/1/organizations/cour-des-comptes/"
  },
}
```

uData – API REST – JSON – partie 2

```
"resources": [  
  {  
    "checksum": null,  
    "created_at": "2022-03-21T11:09:06.693000",  
    "description": null,  
    "filetype": "remote",  
    "format": "csv/utf8",  
    "last_modified": "2022-03-21T11:09:06.693000",  
    "latest": "https://www.data.gouv.fr/fr/datasets/r/c48182e2-a124-4302-899a-de6a92ebc976",  
    "metrics": {"views": 1},  
    "mime": null,  
    "preview_url": null,  
    "published": "2022-03-21T11:09:06.694000",  
    "schema": {},  
    "title": " Les universités à l'horizon 2030 : plus de libertés, plus de responsabilités",  
    "type": "main",  
    "url": "https://www.ccomptes.fr/sites/default/files/2022-03/20211021-NS-Universites-donnees.zip"  
    [ ... ]  
  }  
],  
"slug": "les-universites-a-lhorizon-2030-plus-de-libertes-plus-de-responsabilites",  
"spatial": {  
  "geom": null,  
  "granularity": "country",  
  "zones": ["country:fr"]  
},  
"tags": ["enseignement-superieur", "universites"],  
"temporal_coverage": null,  
"title": "Les universités à l'horizon 2030 : plus de libertés, plus de responsabilités",  
"uri": "https://www.data.gouv.fr/api/1/datasets/les-universites-a-lhorizon-2030-plus-de-libertes-plus-de-responsabilites/"
```

Qu'en retenir ?

- Les métadonnées DCAT proposées par les solutions sont souvent moins riches que les API spécifiques mises en place par les éditeurs même si en théorie son expressivité est très intéressante
- Manque d'homogénéité du renseignement des métadonnées (pas une spécificité des catalogue open data)
- Pas de normalisation de certains champs : granularité spatiale par exemple.
- Pour certaines API, nécessité d'exécuter plusieurs requêtes pour récupérer l'intégralité des informations utiles
- **Même si DCAT paraissait un bon choix au départ de l'étude, c'est un choix trop risqué en l'état actuel des implémentations**

Points de moissonnage des solutions étudiées et retenus (en gras)

<i>Solution</i>	<i>Point de moissonnage possibles</i>	
ArcGIS Online	CSW DCAT , DCAT-AP API REST	
CKAN	CSW DCAT, DCAT-AP	
GeoNetwork	CSW OAI OpenSearch ATOM	DCAT via API REST GeoNetwork Autres API REST GeoNetwork OGC API - records API Elasticsearch
Isogeo	CSW DCAT	
OpenDataSoft	CSW DCAT, DCAT-AP, INSPIRE API REST	
uData (data.gouv.fr)	DCAT API REST	

De quels champs de métadonnées
Géo2France a besoin ?

**Métamodèle de
métadonnées de données
minimal idéal
(selon Géo2France)**

Titre / libellé

Description / résumé

Généalogie

État / Statut de la ressource (withdrawn, obsolete, deprecated, final, onGoing...)

Mots-clefs / thèmes

Auteur / propriétaire / point de contact : libellé, logo, adresse mél

Localisation : emprise, nom du territoire

Granularité spatiale : information structurante pour les données géostatistiques (découpage à la commune, à l'îlot, à l'EPCI, au bassin d'emploi...)

Actualité, information temporelle

Droits d'accès et de réutilisation : licence, mentions à faire figurer sur les réutilisation

Données et ressources associées : URL, protocole, format, libellé, description

Lien vers la fiche de métadonnées d'origine

Illustration graphique

**Spécificités géo
(en discussion entre
Géo2France et
DataGrandEst)**

Identifiant unique

Thème INSPIRE

Type de données (vector...)

Système de coordonnées

Résolution spatiale : distance, échelle, granularité

Absences remarquables

Dictionnaire d'attributs

- pénible à gérer avec ISO 19115 :2005 (dictionnaire d'attributs forcément dans un doc externe à la fiche de métadonnées)
- peut-être plus intéressant avec ISO 19115:2014 (possibilité d'intégrer le dictionnaire d'attribut dans la fiche de métadonnées des données)
- peu satisfaisant avec StatDCAT (avis personnel)
- plus intéressant de présenter les données sous forme tabulaire que de présenter un dictionnaire d'attributs.

Version de la ressource

Langue de la ressource

Quelles difficultés ?

Mise en valeur du propriétaire des données

- Objectif d'une plateforme partenarial :
Mise en avant du producteur des données ou du titulaire des droits de propriété intellectuelle
- Aujourd'hui la plupart des métadonnées et des catalogues mettent en avant : le catalogue d'origine des métadonnées ou l'organisation déclarée en point de contact ou comme « publisher » (voir exemples ArcGIS Online, OpenDataSoft et uData). On a souvent une réappropriation des données au détriment du réel producteur.
- Difficulté à harmoniser les noms des producteurs des données
- Difficulté à afficher un logo pour le producteur s'il n'est pas présent dans l'annuaire du catalogue. Udata propose deux attributs : *logo_thumbnail* et *logo*.

Granularité géographique des données

- Granularité des données :
 - information supportée par data.gouv et OpenDataSoft
 - exemples utilisés par data.gouv :
 - country, fr:commune, fr:collectivite, fr:epci, fr:iris, poi
 - exemples utilisés par OpenDataSoft (maille-geographique) :
 - National, Région, Département, EPCI, Commune, IRIS
- Très utile pour caractériser des données mises à disposition via un outil comme Géoclip
- Usage de cette notion quasi-inexistante avec ISO 19115
- Pratique actuelle pour Géo2France : indication de la granularité à l'aide d'un mot-clef. (moyennement satisfaisant)

Incomplétude des sources de données pour certains catalogue Open Data

- Difficulté en fonction des catalogues :
L'intégralité des liens de téléchargement / API d'accès aux données pas systématiquement récupérables en une seule requête :
Exemple d'OpenDataSoft, requête principale à l'API ne présente pas les liens de téléchargement mais certaines requêtes spécialisées fournissent ces infos. Exemple :
[https://metropole-europeenne-de-lille.opendatasoft.com/\[...\]](https://metropole-europeenne-de-lille.opendatasoft.com/[...])
- Impossibilité avec le moissonneur « harvester-simpleurl » de GeoNetwork de combiner le résultat de plusieurs requêtes.
- Nécessité, pour compenser, d'injecter dans la fiche de métadonnées des informations déduites d'autres informations (URL des liens de téléchargement construites à partir de l'URL du catalogue)

Formats associés aux liens de téléchargement

- Pas de format renseigné pour les liens de distribution dans les fiches ISO 19115:2005 (contrairement à ISO 19115:2014 et à DCAT).
- Ce format est une information importante pour l'utilisateur final qui s'apprête à cliquer sur un lien de téléchargement.

PNG	dreal:dreal_communes_potentiel_radon.png EPSG:2154			
JPEG	dreal:dreal_communes_potentiel_radon.jpg EPSG:2154			
GML	dreal:dreal_communes_potentiel_radon.gml EPSG:2154			
GeoJSON	dreal:dreal_communes_potentiel_radon.json EPSG:2154			

- Nécessité d'ajouter de l'intelligence dans l'interface utilisateur du catalogue pour déduire le format de données associé à chaque lien :
 - analyse de la structure des URL (recherche d'un « format=... » par exemple)
 - test de l'URL

Quels résultats ?

Moissonneurs open data

Transformations XSLT pour convertir les catalogues open data JSON en ISO 19115:2014 :

- ArcGIS Online :
<https://github.com/geonetwork/core-geonetwork/blob/main/schemas/iso19115-3.2018/src/main/plugin/iso19115-3.2018/convert/fromJsonLdEsri.xsl>
- OpenDataSoft :
<https://github.com/geonetwork/core-geonetwork/blob/main/schemas/iso19115-3.2018/src/main/plugin/iso19115-3.2018/convert/fromJsonOpenDataSoft.xsl>

Travail initié par François-Xavier Prunayre (Titellus) et adapté par Camptocamp.

Utilisation d'ISO 19115 comme modèle pivot

Deux raisons :

- Implémentations de DCAT et GeoDACT-AP pas assez matures (y compris dans GeoNetwork)
- Facilité dans le cadre d'un patrimoine existant basé sur ISO 19115 (y compris dans le contexte de projets transfrontaliers – avec la Belgique en l'occurrence)

Interface de découverte
simple et homogène quel
que soit l'origine des
métadonnées

<

Consommation annuelle d'électricité et gaz par département et par secteur d'activité

À PROPOS APERÇU ACCÈS LIENS

À propos

Permet de visualiser l'évolution de 2011 à 2020 des consommations d'électricité et de gaz par secteur d'activité (résidentiel, tertiaire, industriel, agricole ou non affecté) et par département.

Ces données sont publiées dans le respect des règles relatives à la protection des Informations Commercialement Sensibles.
Une notice méthodologique est disponible en PJ.

Les données sont ventilées sur le référentiel INSEE au 1er janvier 2020.

Mots clés

industrie agriculture tertiaire résidentiel secteur d'activité article 179

consommation

Détails

Fiche de métadonnées d'origine

<https://opendata.agenceore.fr/explore/dataset/conso-elec-gaz-annuelle-par-secteur-dactivite-agreee-departement/information/>

CONTACT

Agence ORE & Gestionnaires de réseaux électricité et gaz

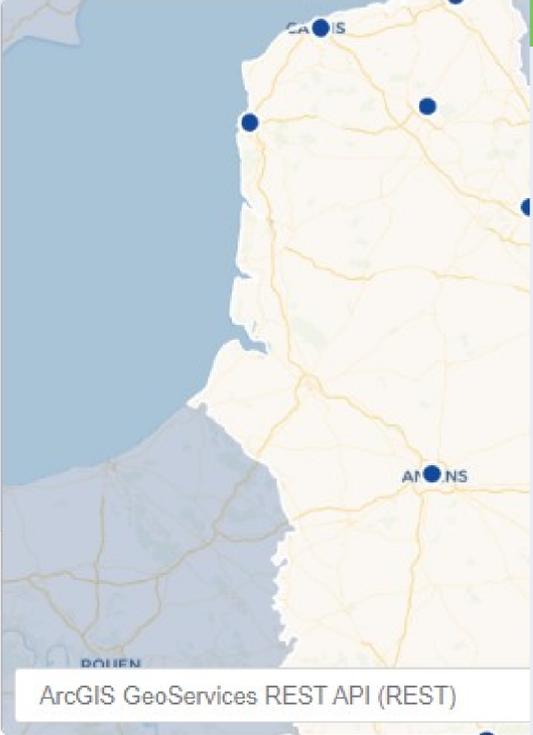
CATALOGUE

Interface de découverte exploitant les liens de téléchargement

ind hdf agglo

À PROPOS APERÇU ACCÈS LIENS

Aperçu VUE CARTE VUE TABLEAU



ArcGIS GeoServices REST API (REST)

Téléchargements

GeoJSON
Geojson

CSV
Csv

ind hdf agglo

À PROPOS APERÇU ACCÈS LIENS

Aperçu VUE CARTE VUE TABLEAU

CSV (csv)

id	X	Y	date_ech	valeur	source	qualif	couleur	lib_zone
1	207996.0157676	6612807.4664509	2020/01/01 00:00:00+00	6	Atmo HDF	Médiocre	#FFAA00	Calais
2	342672.562119118	6512347.09158884	2020/01/01 00:00:00+00	4	Atmo HDF	Bon	#99E600	Douai-Lens
3	365014.383921328	6419907.29480295	2020/01/01 00:00:00+00	4	Atmo HDF	Bon	#99E600	Saint-Quentin
4	308145.708854676	6496529.06879057	2020/01/01 00:00:00+00	4	Atmo HDF	Bon	#99E600	Arras
5	339168.224548946	6556262.08622229	2020/01/01 00:00:00+00	4	Atmo HDF	Bon	#99E600	Lille
6	440968.785684487	6495501.15758161	2020/01/01 00:00:00+00	4	Atmo HDF	Bon	#99E600	Maubeuge
7	276533.199859203	6318360.23321337	2020/01/01 00:00:00+00	5	Atmo HDF	Moyen	#FFFF00	Creil
8	254308.263522326	6428768.19464686	2020/01/01 00:00:00+00	5	Atmo HDF	Moyen	#FFFF00	Amiens
9	252130.854282409	6580252.15296222	2020/01/01 00:00:00+00	4	Atmo HDF	Bon	#99E600	Saint-Omer

4799 enregistrements dans ce jeu de données.

Téléchargements

GeoJSON
Geojson

CSV
Csv

Merci de votre attention

Annexes sur les points de moissonnage

Accès aux métadonnées de GeoNetwork - 1

<i>API exposant les métadonnées</i>	<i>Commentaire</i>
CSW	Interface de moissonnage la plus utilisée par les plateformes de données géo (notamment dans le cadre d'INSPIRE).
OAI	Très peu utilisé. Non disponible avec GeoNetwork 4 au moment de l'étude.
OpenSearch	N'est plus utilisé.
ATOM	N'est plus utilisé. Non disponible avec GeoNetwork 4 au moment de l'étude.
API Elasticsearch	Utiliser l'API d'Elasticsearch intégré à GeoNetwork 4 est également possible.

Accès aux métadonnées de GeoNetwork - 2

<i>API exposant les métadonnées</i>	<i>Commentaire</i>
DCAT via API REST GeoNetwork	<p>Deux API de GeoNetwork 3 (rdf.search et rdf.metadata.get) proposaient un export DCAT ancien. Ces deux services ne sont plus accessibles avec GeoNetwork 4. Néanmoins, cet export DCAT est toujours utilisé au niveau du service CSW.</p>
Autres API REST GeoNetwork	<p>Les API offrent de nombreuses possibilités mais elles sont spécifiques aux dernières versions de GeoNetwork et peuvent nécessiter une succession d'appels en fonctions des besoins. Exemples d'API utiles :</p> <ul style="list-style-type: none">- records- records/links- search/records (proxy pour recherches Elastic)- related
OGC API - records	<p>Intégré au projet "microservices" de GeoNetwork. Pour l'utiliser il faut déployer un micro-service en parallèle de GeoNetwork. L'un des output formats de cette API est le profil GeoDCAT-AP. Les specs de DCAT et GeoDCAT-AP n'étant pas forcément limpides il existe quelques champs dont les correspondances pourraient être améliorées. Cette implémentation est totalement indépendante du code utilisé pour produire les enregistrements DCAT fournis par le service CSW. Autre format supporté intéressant pour l'indexation par les grands moteurs de recherche : Schema.org / JSON-LD.</p>

Accès aux métadonnées d'ArcGIS Online

<i>API exposant les métadonnées</i>	<i>Commentaire</i>
CSW	<p>CSW n'est pas supporté par ArcGIS Online même s'il est possible d'accéder à des fiches de métadonnées individuelles.</p> <p>ISO 19139 : pas de support des spécificités d'INSPIRE. Le modèle complet d'INSPIRE peut être géré en interne à ArcGIS Enterprise mais, actuellement, il ne peut pas être exposé sur le web.</p>
DCAT	<p>Une version minimale de DCAT implémentée et activable via l'option gratuite Open Data de la solution. Volonté d'ESRI d'enrichir ce modèle pour mieux correspondre aux attendus d'INSPIRE.</p> <p>Exemple de point d'accès DCAT : https://data-atmo-hdf.opendata.arcgis.com/data.json</p> <p>C'est ce point d'accès qui est utilisé aujourd'hui par data.gouv pour moissonner des plateformes ESRI.</p>
DCAT-AP	<p>Support de DCAT-AP via un autre point d'accès (également activable via l'option gratuite Open Data) : https://data-atmo-hdf.opendata.arcgis.com/catalog/dcat-ap/2.0.1.json</p>
OGC API Records	<p>L'implémentation d'OGC API Records fait partie de la feuille de route d'ArcGIS Online pour 2022.</p>
API REST	<p>Les API REST d'ArcGIS peuvent permettre d'accéder aux métadonnées des données. Néanmoins, elles n'ont pas de réelle plus-value dans le cadre du moissonnage de métadonnées.</p>

Accès aux métadonnées d'OpenDataSoft

<i>API exposant les métadonnées</i>	<i>Commentaire</i>
CSW	<p>Exemple de fiche de métadonnées ISO 19115 retournée par OpenDataSoft : https://data.laregion.fr/api/... qui correspond à ce jeu de données : https://data.laregion.fr/explore/dataset/dreal-occitanie-bassins-de-vie-en-occitanie/information/</p> <p>Les métadonnées proposées par cette interface sont moins complètes que celles proposées par les autres API d'OpenDataSoft (en particulier la licence et les liens de téléchargement). Les services WFS internes d'accès aux données sont présents dans la fiche.</p>
DCAT DCAT-AP INSPIRE	<p>OpenDataSoft propose des extensions au modèle de métadonnées interne pour DCAT, DCAT-AP et INSPIRE.</p> <p>Documentation des champs de métadonnées standard Documentation des champs supplémentaires</p>
API REST	<p>API très complète intégrant des possibilités de filtre étendues.</p> <p>Doc pour l'export d'un catalogue : https://help.opendatasoft.com/apis/ods-search-v2/#exporting-datasets</p> <p>Doc pour récupérer le détail pour un jeu de données : https://help.opendatasoft.com/apis/ods-search-v2/#looking-up-a-dataset</p>

Accès aux métadonnées de CKAN

<i>API exposant les métadonnées</i>	<i>Commentaire</i>
CSW	Support de CSW via le plugin ckan-ext-spatial. N'expose que les champs de métadonnées correspondant au schéma interne de CKAN. Les champs personnalisés ne sont pas pris en compte. Néanmoins, le plugin est extensible (développements à réaliser).
DCAT DCAT-AP	Bonne intégration, récupération complète des informations DCAT via l'extension ckanext-dcat. Très bonne documentation : https://github.com/ckan/ckanext-dcat#rdf-dcat-to-ckan-dataset-mapping
API REST	API trop spécifique pour servir de point d'entrée à un moissonnage s'appuyant sur des formats standards.

Accès aux métadonnées d'Isogeo

<i>API exposant les métadonnées</i>	<i>Commentaire</i>
CSW	Intégralité des métadonnées accessible via cette interface.
DCAT	Prévu pour T4 2021

Accès aux métadonnées de uData (data.gouv)

<i>API exposant les métadonnées</i>	<i>Commentaire</i>
DCAT	Format JSON ou XML. Pour l'ensemble du catalogue : https://www.data.gouv.fr/fr/catalog.json Pour une organisation : https://www.data.gouv.fr/fr/organizations/meteo-france/catalog.json Pour un jeu de données : https://www.data.gouv.fr/fr/datasets/projections-climatiques-sur-la-zone-large-molene-sur-un-mois/rdf.json



+33 6 52 45 58 80



contact@bchartier.net



<https://bchartier.net>



<https://www.linkedin.com/in/benjaminchartier>



<https://github.com/bchartier>